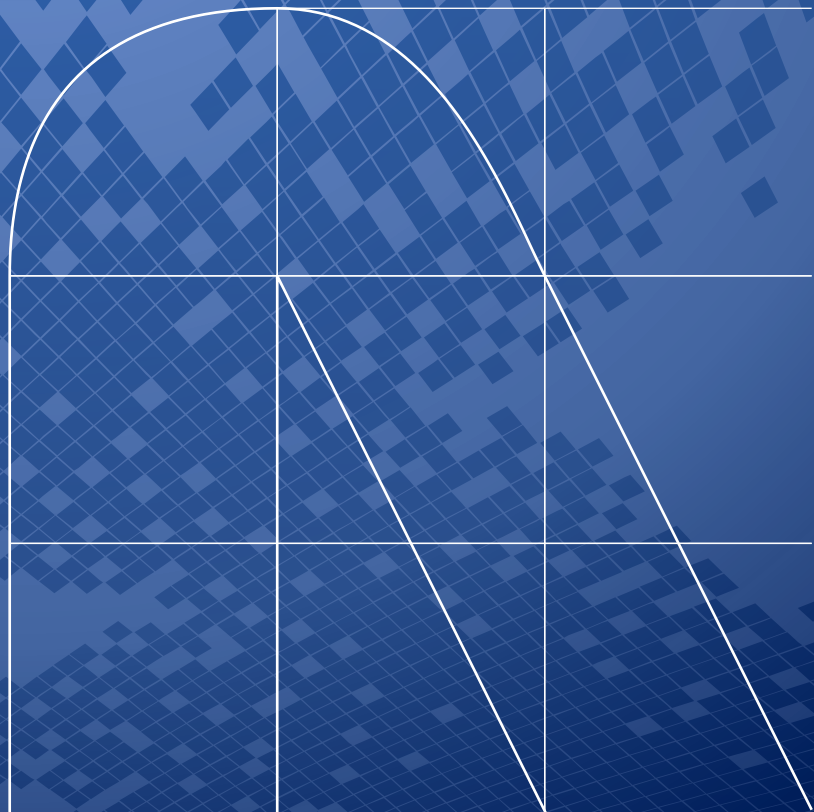# NTT DATA

Generative AI Unveiled: Navigating the Complexities and Opportunities

NTT DATA point of view on Generative AI and its Complexities and Opportunities in real world scenarios.

NTT DATA Point of View

The Hype is Real

ChatGPT Sprints to One Million Users

Time it took for selected online services to reach one million users

Service	Launched	Time
Netflix	1999	3.5 years
Kickstarter*	2009	2.5 years
Airbnb**	2008	2.5 years
Twitter	2006	2 years
Foursquare***	2009	13 months
Facebook	2004	10 months
Dropbox	2008	7 months
Spotify	2008	5 months
Instagram***	2010	2.5 months
ChatGPT	2022	5 days

* one million backers ** one million nights booked *** one million downloads
Source: Company announcements via Business Insider/Linkedin

statista

In the dynamic landscape of 2022, a captivating buzzword took center stage: the rise of GPT, personified by the enigmatic ChatGPT, representing the cutting edge of Generative Artificial Intelligence (GAI). While this technological marvel had quietly existed in the shadows for a little over five years[1], it suddenly exploded onto the scene, captivating the collective imagination, and swiftly establishing itself as one of history's most rapidly embraced innovations.

In 20 years following the Internet space, we cannot recall a faster ramp in a consumer internet app.[2]

Generative AI has cast its spell over a global audience, setting off an exhilarating wave of fresh applications that sparked creativity and innovation to unprecedented heights. From creating art and melodies to crafting text so eerily human-like it was indistinguishable, Generative AI is reshaping our interaction with machines, unlocking a new dimension of possibilities. With its uncanny ability to mimic human conversations and navigate critical decision-making, Generative AI offers exciting glimpses of its profound disruptive potential.

Yet, amid this whirlwind of progress, it's crucial to recognize that Generative AI remains in its infancy, concealing vast realms of uncharted territory. As we stand on the face of an era where humans and intelligent machines join forces to accomplish feats beyond individual reach, Generative AI is poised to redefine the very nature of work. It's destined to become an essential companion in various aspects of our daily lives, sometimes operating invisibly in the background.
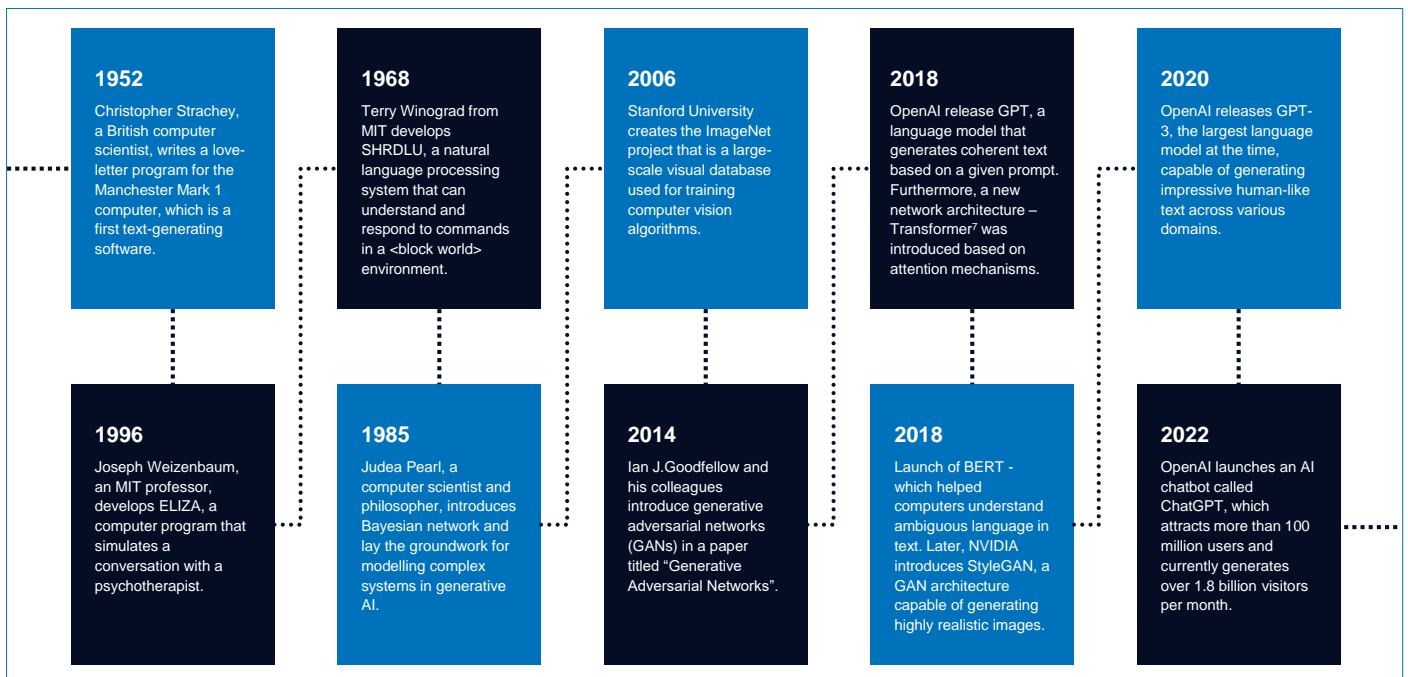
Nevertheless, we embark on a thrilling journey into the heart of Generative AI, presenting a comprehensive guide for beginners. But beware, for along the path of discovery lie three treacherous mistakes that, if left unheeded, could haunt even the most daring pioneers in the world of Generative AI.

According to Gartner®, "Generative artificial intelligence (AI) is positioned on the Peak of Inflated Expectations on the Gartner, Inc. Hype Cycle for Emerging Technologies, 2023, projected to reach transformational benefit within two to five years.[3]

The Evolution of Generative AI

The evolution of Generative AI has been nothing short of remarkable in recent years, it is so quick and so accurate that it almost feels like sorcery. The quality and capabilities of these AI have undergone a profound transformation from the early basic experiments to real world applications. Currently, there are hardly any fields or sectors were, Generative AI has not left a mark, let it be healthcare to entertainment, content creation or data analysis.

As we stand on the edge of future, where the possibilities for Generative AI are boundless. With continuous advancement in machine learning and deep neural network, it promises to redefine how we interact with technology, right from chatbot to creative content generation. The journey of Gen AI is very fascinating, and the surprising fact is the journey of AI started almost 50 years ago.

Timeline

1952
Christopher Strachey, a British computer scientist, writes a love-letter program for the Manchester Mark 1 computer, which is a first text-generating software.

1968
Terry Winograd from MIT develops SHRDLU, a natural language processing system that can understand and respond to commands in a <block world> environment.

2006
Stanford University creates the ImageNet project that is a large-scale visual database used for training computer vision algorithms.

2018
OpenAI release GPT, a language model that generates coherent text based on a given prompt. Furthermore, a new network architecture – Transformer[7] was introduced based on attention mechanisms.

2020
OpenAI releases GPT-3, the largest language model at the time, capable of generating impressive human-like text across various domains.

1996
Joseph Weizenbaum, an MIT professor, develops ELIZA, a computer program that simulates a conversation with a psychotherapist.

1985
Judea Pearl, a computer scientist and philosopher, introduces Bayesian network and lay the groundwork for modelling complex systems in generative AI.

2014
Ian J.Goodfellow and his colleagues introduce generative adversarial networks (GANs) in a paper titled "Generative Adversarial Networks".

2018
Launch of BERT - which helped computers understand ambiguous language in text. Later, NVIDIA introduces StyleGAN, a GAN architecture capable of generating highly realistic images.

2022
OpenAI launches an AI chatbot called ChatGPT, which attracts more than 100 million users and currently generates over 1.8 billion visitors per month.

Demystifying Generative AI

Generative AI, at its core, is a subfield of artificial intelligence that focuses on creating algorithms and models capable of generating data that resembles human-created content. These algorithms leverage neural networks, with notable approaches like Variational Autoencoders (VAEs) for structured data, Generative Adversarial Networks (GANs) for highly realistic content generation, and recurrent neural networks (RNNs) or Transformers for sequential data, such as text or speech. The training process involves exposing these models to extensive datasets, allowing them to learn patterns and correlations within the data. However, challenges such as mode collapse in GANs and ensuring high-quality training data are crucial considerations in the development and ethical deployment of Generative AI.

Generative AI finds applications in diverse fields, from natural language generation for chatbots and content creation to image synthesis for art and design. By demystifying Generative AI and understanding its underlying technical principles,

businesses and researchers can harness its potential to automate content generation, improve creative processes, and innovate across various domains while remaining vigilant about ethical considerations and data quality.
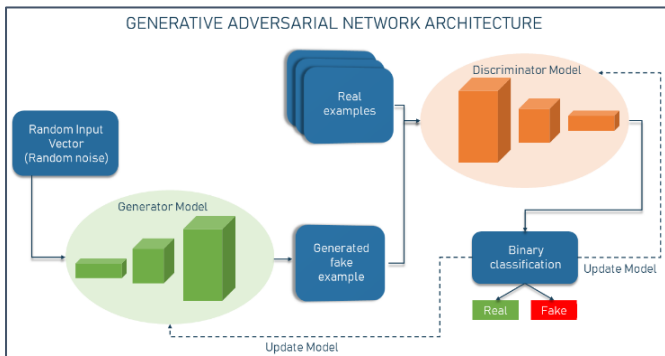
Imagine Generative AI as a master chef of data creation. Just as a chef can craft an endless variety of dishes, Generative AI can produce a wide range of content. Think of Variational Autoencoders (VAEs) as the chef's secret recipe book. VAEs allow the chef to understand the essence of various cuisines and create new, unique dishes by blending flavors creatively. In contrast, Generative Adversarial Networks (GANs) act like a culinary competition. Here, the chef (generator) strives to cook dishes that are indistinguishable from the real thing, while a food critic (discriminator) tries to tell the difference. This ongoing rivalry results in the chef continually improving and creating incredibly lifelike meals.

Are you watching closely?

The Pledge

A foundational training processes, where the model learns from a massive dataset, becoming well-versed in language and knowledge.

The Turn
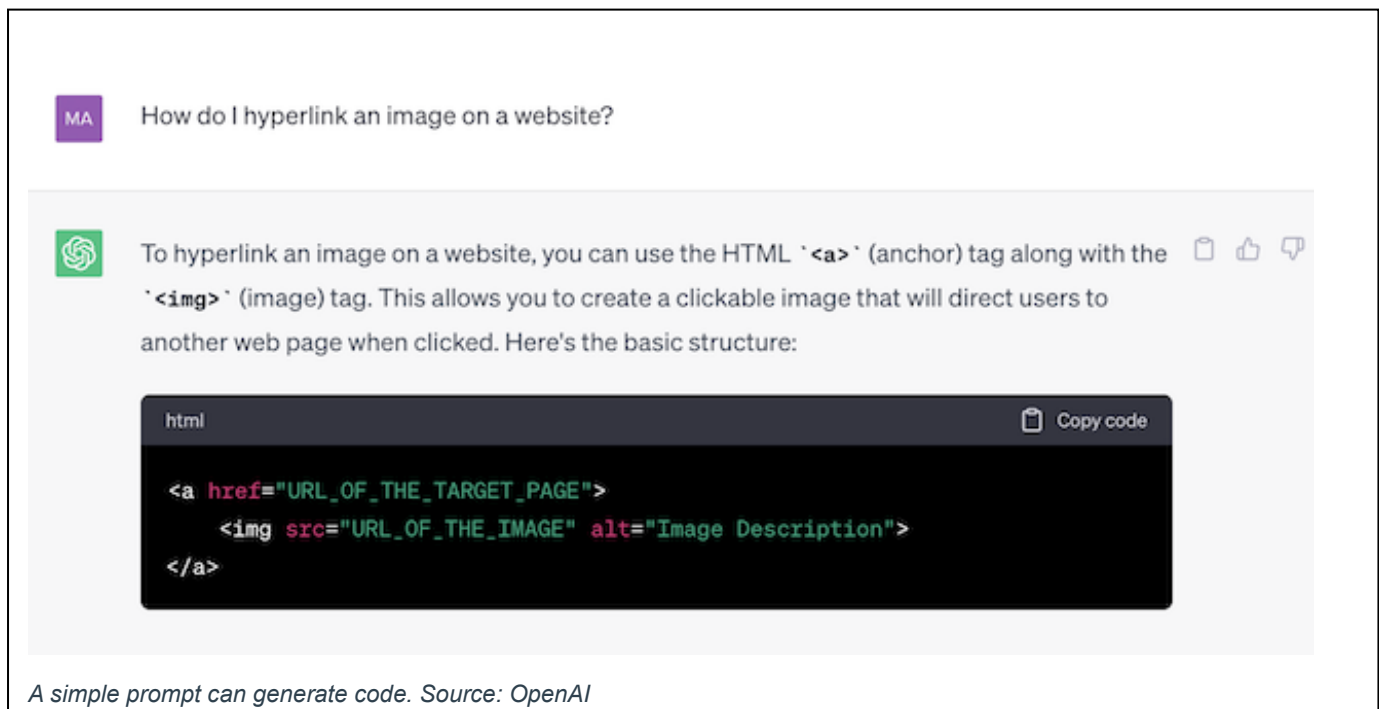
The ability to take ordinary inputs (user queries) and transform them in to complex, contextually relevant outputs. It's the magic that happens when it generates coherent and engaging responses.

The Prestige

This is the moment when the AI produces a response that's almost indistinguishable from human-generated content. It's the awe-inspiring result of its training, processing, and creative capabilities.

Where, How and Why?

A simple prompt can generate code. Source: OpenAI

Let's embark on a journey of transformation, reshaping the way we approach tasks, with Generative AI as our trusted co-pilots. By integrating this technology into our operations, companies can unlock a realm of endless possibilities. It's a catalyst that not only amplifies tasks but also opens new doors of opportunity.

Here's a glimpse of where, how, and why Generative AI is poised to revolutionize our world:

Conversations: Picture having Generative AI as your co-pilot in customer interactions. This smart, cognitive bot understands your customers' intent and crafts responses that elevate the quality of their experience. It's so advanced that the conversations it generates are nearly indistinguishable from human interaction. For Example: ChatGPT can aid medical professionals in healthcare by furnishing real-time patient data, proposing treatment alternatives, and responding to medical questions in easy-to-understand language, thus enhancing patient understanding and ultimately improving patient care and efficiency in healthcare settings.

NTT DATA offers specialized services for Public Administrators for easy understanding of complex regulations and similar services for financial sector clients for easy understanding of complex financial topics.

Content Creation: Generative AI brings in revolutionary capabilities to empower editors, writers, and content creators. It has the potential to truly transform how content is created across various domains and instead of displacing knowledge workers and content writers, Generative AI must be leveraged to enhance and boost their capabilities. It can generate human like text, enabling automatic draft of articles, emails and even poems. This creativity also extends to visual content as well, with tools like DALL.E, producing unique and imaginative images based on textual input. It does not stop here, Gen AI tools can compose music with AI composers, offering a modern and creative approach to artistic expression.

Knowledge Management: Knowledge management had its own set limitations in handling extensive datasets and extracting meaningful insights as traditional methods relied on manual curation and keyword-based search which also leads to gaps in information retrieval, and at times omitting desired information. With the ability to handle large datasets, Generative AI plays a pivotal role in knowledge management. Transformer based models (LLMs), demonstrate proficiency in grasping contextual nuances. This technology helps in enhancing search engine capabilities and provide more accurate context aware results. Moreover, these transformative models also help in summarizing and condensing lengthy content into succinct yet informative.

Integration of such tools help us harness information effectively.

Content Discovery: Gen AI can shape content discovery through personalized recommendations can providing tailored content based on user behavior. Additionally, with advanced natural language processing, Generative AI refines content understanding, improving categorization and tagging for more efficient search algorithms. In essence, generative models elevate content discovery by offering a nuanced and personalized experience across various platforms.

NTT's Digital human[6] was showcased at Open Championship 2023, which took place at Royal Liverpool.
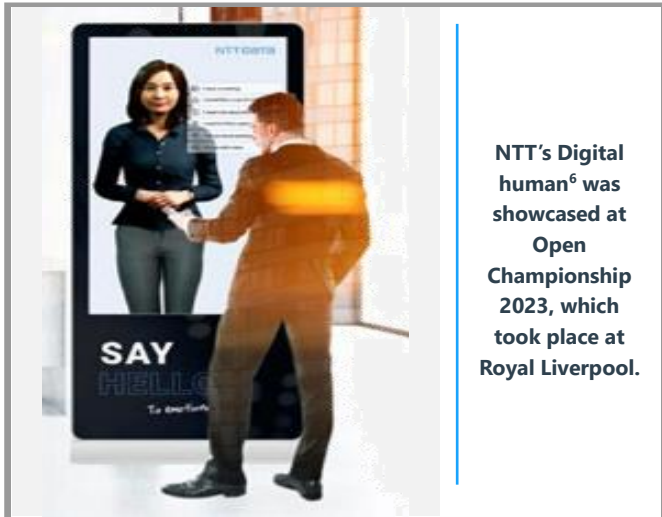
In a recent collaboration NTT DATA implemented its conversational AI Eva with GenAI to reimagine CX for a French personal care and cosmetics giant.

Process Automation: The era of business process automation is upon us. Generative AI, armed with extensive contextual knowledge and historical data, streamlines, and accelerates customer interactions by automatically analyzing processes to be automated, reducing manual effort. We have a strong experience and value offerings such as:

- Dolffia, which is a multi-cloud Generative AI solution to empower your business by optimizing document processing, enhancing content creation, and accelerating data insights for tailored and scalable industry-specific and cross-functional use cases.
- LITRON®, a document comprehension AI that can read and comprehend Japanese texts quickly and accurately. It features the LITRON® Generative Assistant, which combines LLM with LITRON® Generative

Assistant technology to enable searching of internal documents and providing responses in a chat format. This advanced technology empowers organizations to unlock the full potential of their Japanese language documents and maximize the value derived from them.

Software Development and Maintenance: Generative AI is proficient in several popular programing languages and frameworks and can seamlessly convert between them. Furthermore, it can comprehend human-written problem statements and instantly generate the required code, keeping pace with the latest advancements and security considerations. GitHub Copilot is an example of Generative AI used in coding applications. This copilot serves as an intelligent code completion tool, which assists developer's snippet of codes. This feat is achieved by understanding the context and generating code in real-time, which significantly increases productivity.

Art and Creativity: Generative AI taps into real-time global trends and knowledge to assist creators in generating engaging content with a simple click. Whether it's a captivating picture or a mesmerizing piece of music, Generative AI can do it all. DALL.E and MidJourney are examples of Generative AI models dedicated to artistic endeavors.

DALL.E, for instance, can generate unique and visually stunning digital artworks, offering a fresh perspective on creativity in the art world. MidJourney, on the other hand, helps artists generate new ideas, refine their craft, and create more impactful works.

Road to Implementation: Integrating Generative AI in Businesses.

Before we dive into steps of incorporating Generative AI into corporate operations, it is extremely essential to understand that it is more than technical undertaking. It is a strategic step, which needs precision, alignment with organizational goals and needs continuous monitoring and refinement.

Let's list a few key steps that would paint a picture of implementation of Generative AI.

Step 1: Define Clear Business Objectives and assess Impactful Use Cases

Begin by articulating well-defined business goals that align with the overarching organizational vision. This foundational step not only shapes the contours of Generative AI integration but also provides a focused lens to identify specific areas within the organization where this transformative technology can be effectively applied.

We also need to thoroughly evaluate potential use cases for Generative AI, taking into consideration factors such as implementation ease, projected return on investment (ROI), and their direct correlation with established business objectives. This meticulous evaluation ensures that the adoption of Generative AI is not just a technological advancement but a strategic move that significantly impacts organizational outcomes.

Step 2: Project Discovery & Precision Planning

In this step we not only need to identify the technical challenges (distinct from broader business issues), but also put together technical solutions, including model choices, fine-tuning strategies, and integration approaches with external knowledge bases along with appropriate technology stack, encompassing cloud services, frameworks, libraries, and vector databases.

It is equally important to establish a technical as well as non-technical success metrics and conduct a comprehensive cost assessment. Prompt engineering plays a vital role in achieving efficient AI systems and must be actively monitored to reduce unnecessary computations and to optimize resource utilization.

Step 3: Proof of Concept & Pilot Implementation

Prior to large scale deployment, it is essential to conduct a meticulous proof of concept on a small scale to validate its feasibility (in a controlled environment). This stage serves as an early identification mechanism for challenges and allows for refinement.

A user-friendly interface (UI) integrated with existing systems may help in active solicitation of feedback for ongoing improvement, which assists in further fine-tuning and optimization. In the excitement, one must not forget the compliance with industry regulations and data privacy laws.

Step 4: Full Implementation - Realizing the Vision & Optimization/Continuous Maintenance

Based on the validated pilot and the inputs received in the previous steps, we can now execute a full-scale deployment of Generative AI. This phase involves not just technological implementation but a strategic incorporation of the refined model into the organizational fabric.

It requires a perpetual optimization, maintenance, and governance cycle post-implementation. To keep the refining the tool as per the evolving organizational needs.

It is essential to define a reference model and a standardized enterprise architecture/framework, that provides a blueprint for desired outcome and can ensure critical success factors when going from Proof of Concept (PoC) to Production while enabling scalability, performance, and keeping an eye on security, privacy and governance aspects.
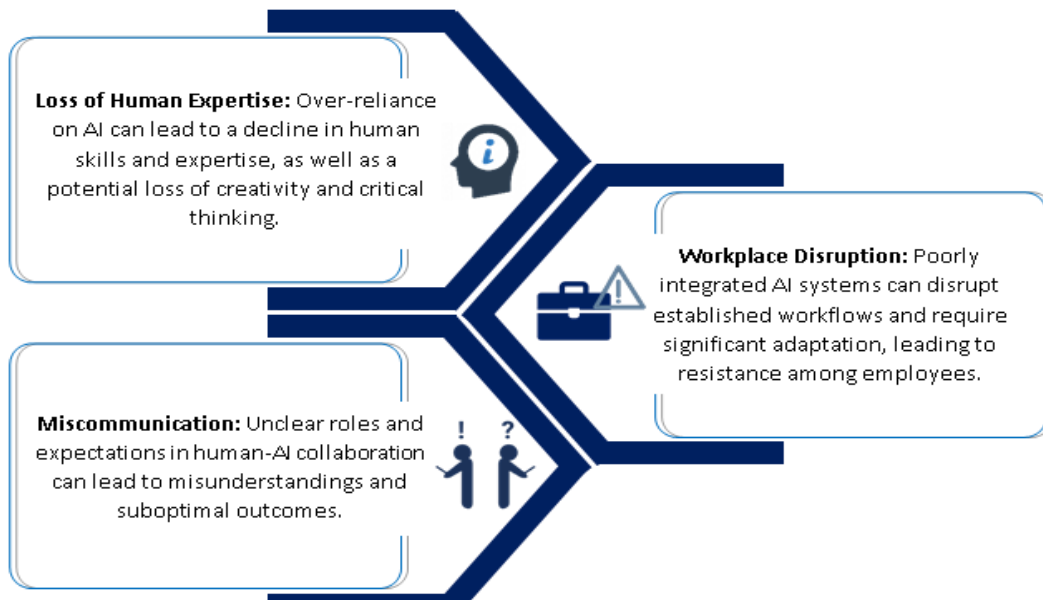
Mistakes that Might Haunt you Later.

As we embrace the transformative potential of Gen AI, it's crucial to acknowledge the challenges that can be inadvertently overlooked during its implementation. Despite the numerous advantages Gen AI offers, many businesses remain hesitant to fully embrace it.

One primary reason for this hesitancy lies in the complexity of Gen AI systems. There have been Black box issues in AI, this inability for us to see how deep learning systems make their decisions. In addition, concerns regarding scalability, compatibility with existing systems, and the need for specialized talent can lead to hesitancy.

Moreover, ethical considerations are at the forefront of these reservations. Ensuring transparency, accountability, data privacy, and the potential for harmonious human-AI collaboration are paramount.

Loss of Human Expertise: Over-reliance on AI can lead to a decline in human skills and expertise, as well as a potential loss of creativity and critical thinking.

Workplace Disruption: Poorly integrated AI systems can disrupt established workflows and require significant adaptation, leading to resistance among employees.

Miscommunication: Unclear roles and expectations in human-AI collaboration can lead to misunderstandings and suboptimal outcomes.

Failure to address these critical issues could lead to three haunting mistakes:

Data Privacy and Security:

Data privacy and security are paramount in Gen AI implementation, as AI models often require access to sensitive information. Failing to prioritize data privacy and security can lead to serious repercussions:

To avoid these pitfalls, organizations must implement robust data encryption, access controls, and data anonymization techniques. Regular security audits and compliance with relevant data protection laws are crucial.

In order to comply with regulatory frameworks out there such as GDPR of EU and CCPA of California, USA, it is essential to establish that who has control over the data (data sovereignty) and where the data is stored (data residency), which adds an additional layer of complexity.

Hence, one must conduct regular security audits to ensure compliance.

Human AI Collaboration:

Efficient collaboration between humans and Gen AI is essential for productive and effective workflows. Neglecting to strike the right balance can lead to inefficiencies and frustration:

To mitigate these challenges, organizations should carefully design workflows that incorporate Gen AI as a supportive tool rather than a replacement. Training and upskilling employees in

AI-related skills can also foster more effective human-AI collaboration.

Measuring Efficiency and Return on AI Investment (ROAI):

Although Generative AI promises unprecedented productivity, it is essential to measure the Gen AI efficiency and its impact. Hence, it is important to establish KPIs and validate them against business goals to deduce Return on AI Investment (ROAI). Calculating ROAI not only provides technical understanding, but also economic value generated. This analysis helps the orgnization with long term benefits.

Sample Selection for PoC and Adopting Right Framework:

The transition from PoC to Production requires an appropriate sample for testing critical applications under Generative AI solutions. Hence, it is important to pick samples that would be relevant in real world scenarios. Also, post transition adopting the right framework is essential for integrating Generative AI into existing IT systems.

Navigating the Boundaries of Gen AI: Understanding its Limitations

Hallucination

Gen AI models are susceptible to generating content that may appear imaginative but lacks grounding in reality

Bias

AI models can perpetuate these biases, leading to discriminatory or unfair outputs.

Lack of Human Reasoning

Generative AI systems are based on statistical features. However, it lacks the ability for true human reasoning.

Limited Context Window

Current models have a context window of a few thousand words, it may struggle with providing comprehensive context awareness.

While Generative AI (Gen AI) continues to astonish us with its creative prowess, it's crucial to recognize that even the most brilliant artists have their constraints. In our exploration of the Gen AI landscape, let's venture into the intriguing realm of its limitations, where the digital brush sometimes paints outside the lines. Hence, it's imperative to acknowledge that while this technology holds immense promise, it is not without its limitations. These limitations are like the intriguing imperfections in a masterpiece, reminding us of the need for discernment and careful consideration.

Hallucinations:

Much like dreams can take unexpected and nonsensical turns, Gen AI models are susceptible to generating content that may appear imaginative but lacks grounding in reality. These AI "hallucinations" can result in outputs that, while creative, are factually incorrect or illogical. This limitation underscores the need for human validation and context-awareness when interpreting Gen AI-generated content.

Bias:

Gen AI learns from vast datasets, which may inadvertently contain biases present in society. Consequently, AI models can perpetuate these biases, leading to discriminatory or unfair outputs. Recognizing and mitigating bias remains a significant challenge in the development and deployment of Gen AI systems, necessitating vigilant oversight and ethical considerations.

Lack of Human Reasoning:

While Gen AI excels at pattern recognition and content generation, it lacks the ability for true human reasoning. It cannot comprehend concepts, make judgments based on ethical or moral principles, or engage in complex decision-making as humans do. This limitation underscores the importance of human supervision and guidance when using Gen AI in critical applications.

Limited Context Overview:

Gen AI operates within the constraints of the data it has been trained on. It may struggle with providing comprehensive context awareness, especially in situations where the information required extends beyond its training data. This can result in responses that lack nuance and fail to consider broader contextual factors.

Understanding and navigating these limitations is essential for harnessing the potential of Gen AI effectively. By acknowledging these boundaries, organizations can make informed decisions about the use of Gen AI and ensure that its applications align with both ethical and practical considerations. In doing so, they can unlock the true value of this transformative technology while mitigating its inherent challenges. Nevertheless, this is being addressed now with Retrieval Augmented Generation (RAG) approach, which can parse relevant information from a large corpus of text, which enables it to generate high-quality text that is both coherent and diverse.

This leads to more accurate and informative responses, making it ideal for applications such as text summarization, machine translation, chatbots, and content creation.

Economic Impact

Generative AI's impact on productivity could add trillions of dollars in value to the global economy

About 75 percent of the value that Generative AI use cases could deliver falls across four areas: Customer operations, marketing and sales, software engineering, and R&D

Generative AI will have a significant impact across all industry sectors and could deliver value equal to an additional $200 billion to $340 billion annually

Generative AI has the potential to change the anatomy of work, augmenting the capabilities of individual workers by automating some of their individual activities.

Generative AI can substantially increase labour productivity across the economy, but that will require investments to support workers as they shift work activities or change jobs.

Source: Mckinsey[4]

Generative AI at NTT DATA

NTT DATA has been conducting research & development of services and products utilizing technologies focused on natural language processing research and groundbreaking models such as GPT since 2018, including publishing white papers about LLMs (including GPT-2 and GPT-3) in 2021[1]. We are continuing our journey of development with foundational Japanese language models, which laid the groundwork for our continued innovation. In November 2020, we proudly introduced a groundbreaking free large language model, boasting an impressive 1.6 billion parameters. This model was meticulously trained using an extensive corpus of web dialogue data and high-quality dialogue data that emerged from our extensive research into Japanese dialogue systems.

Our vision is firmly rooted in the belief that Generative AI, particularly large language models, will continue to shape the future of the AI landscape. Our longstanding commitment to natural language processing has uniquely positioned us to seize the burgeoning demand for Generative AI as an opportunity to leverage our technological expertise.

To help our AI offerings scalability, efficiency, and adoption time reduction, we have developed our Asset-Based Consulting (ABC) strategy. This approach enables us to offer the best scalable services and fast implementation to our clients in combination with our top talent in various offerings:

IDP & Knowledge Management

Meeting the needs of our customers in areas such as Intelligent Document Processing (IDP), knowledge management, and process automation.

Conversational AI

Our solutions leverage the code generation capabilities of foundational models to automate several tasks in the end-2-end of software dev and maintenance as well.

Code Generation and Transformation

Leveraging Generative AI, we have efficiently modernized legacy apps, improving their performance, scalability, and maintainability.

Intelligent Managed Services

We leverage Generative AI to automate enterprise outcomes, revolutionizing IT environment management through streamlined resource provisioning, workflow optimization, and infrastructure efficiency.

Our mission is to bolster the practical applications of language-based AI by enhancing its fundamental attributes, including efficiency, reliability, ethical considerations, and customizability. Furthermore, we are committed to bridging the gap between language-based AI and the tangible world. We aspire to create AI systems that can manifest real-world outcomes, transcending mere computational algorithms. Our dedication lies in empowering AI to enact positive change in the physical realm, aligning with our vision for a future where technology seamlessly augments our everyday lives.

Embrace the Change: Leveraging Generative AI in the Bigger Picture

Generative AI (Gen AI) has undoubtedly emerged as a transformative force, reshaping industries, and unlocking new possibilities. As we embrace this change, it's vital to recognize that Gen AI is a remarkable tool, but it is not the sole solution in our toolkit. It's more akin to a versatile instrument in an orchestra, complementing the symphony of solutions required to address complex challenges.

While Gen AI holds great promise, it's essential to conduct a cost-benefit analysis. Gen AI's power comes at a price, both in terms of GPU processing and electricity consumption. It's important to assess whether the return on investment (ROI) justifies its utilization, particularly when the task at hand may be more efficiently handled by human expertise or other AI techniques. Infrastructure considerations, such as the need for substantial computing resources, must be weighed against the potential benefits.

Therefore, a smart tweak would be employing smaller LLMs to strike a balance between performance and practicality. This will not only enhance efficiency and transparency but also will be cost effective making them attractive alternatives to larger LLMs for a wide range of applications.

For instance, a media company embracing Gen AI for content generation may discover that it can efficiently produce a high volume of articles. Still, the cost of GPU clusters and electricity may erode the cost savings if not carefully managed.

Here, adoption of smaller LLMs would be cost-effective, efficient, and easier to deploy solution.

Conversely, a financial institution using Gen AI for data analysis may find that the speed and accuracy of AI-driven insights justify the computational expenses.

NTT's 'tsuzumi'[5] is a lightweight model available with 600 million and 7 billion version with top-level Japanese language processing capabilities. It is approximately 300 times lighter than OpenAI's GPT-3 and is claimed to surpass GPT-3.5 and top domestic LLMs as per Rakuda, a benchmark for generative AI.

In this era of rapid technological advancement, embracing change is paramount. Gen AI, with its creative abilities and data-driven insights, can undoubtedly drive innovation and efficiency. However, it should be viewed as a valuable addition to our arsenal of solutions, rather than a one-size-fits-all remedy.

As we journey forward, let's remember that technology, including **Gen AI, is a means to an end, not the end itself**. The true measure of success lies in harnessing these tools effectively to solve real-world problems, while being mindful of ethical considerations and the cost-effectiveness of our choices. By embracing the change thoughtfully and holistically, we can navigate the evolving landscape of AI with confidence and drive innovation for a brighter future.

Let's get started

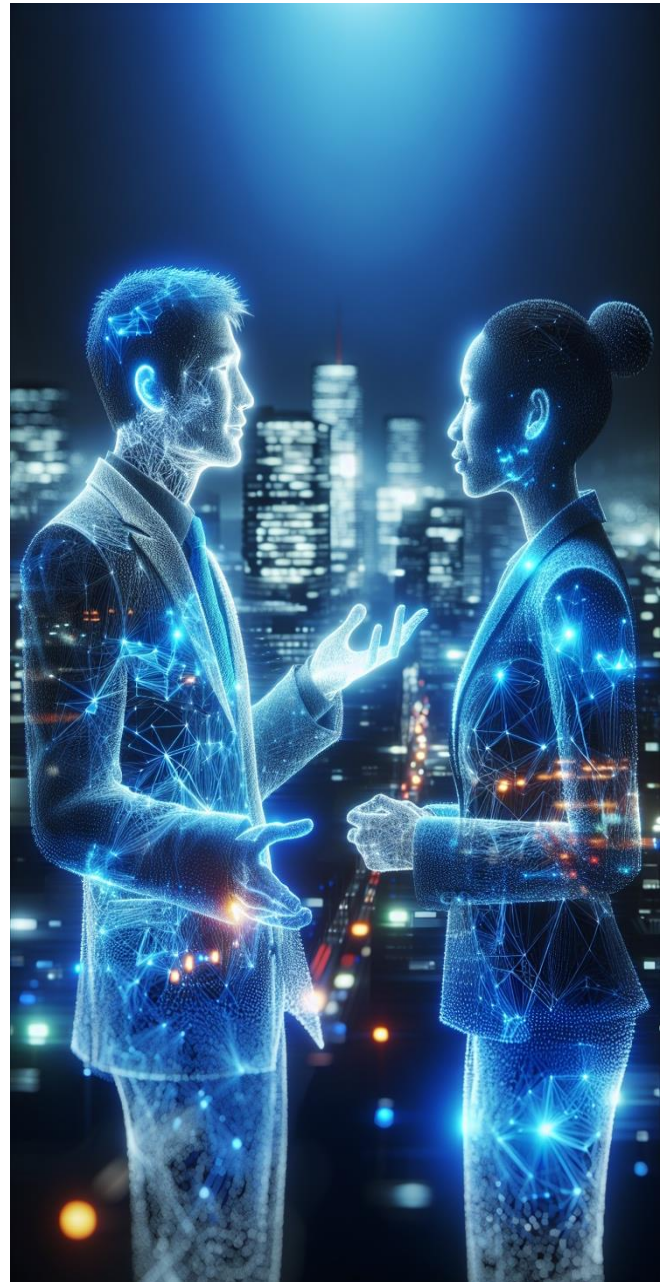**See what NTT DATA can do for you.**

- Deep industry expertise and market-leading technologies
- Tailored capabilities with your objectives in mind
- Partnerships to help you build and realize your vision.

Contact one of our authors or visit **nttdata.com** to learn more.

Sources

1. Languages, models and business applications (nttdata.com)
2. https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/
3. Gartner, Gartner Places Generative AI on the Peak of Inflated Expectations on the 2023 Hype Cycle for Emerging Technologies, https://www.gartner.com/en/newsroom/press-releases/2023-08-16-gartner-places-generative-ai-on-the-peak-of-inflated-expectations-on-the-2023-hype-cycle-for-emerging-technologies, August, 16, 2023. GARTNER is a registered trademark and service mark of Gartner, Inc. and/or its affiliates in the U.S. and internationally and is used herein with permission. All rights reserved.
4. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
5. NTT's Large Language Models "tsuzumi" | NTT R&D Website (rd.ntt)
6. https://www.nttdata.com/global/en/about-us/focus/how-ai-powered-digital-humans-are-revolutionising-customer-experience
7. https://arxiv.org/abs/1706.03762

All images in this report were generated using Azure Open AI

NTT DATA